

What Is Claimed Is:

1. An automated method for sequence evaluation used to compare sequence information relating to at least one sample against sequence information relating to at least one reference, the method comprising:

acquiring sequence information relating to the at least one sample and to the at least one reference;

evaluating the sequence information relating to the at least one sample to identify ambiguous bases present within the sample sequence information by applying a rule-based criteria wherein ambiguous bases are distinguished from unambiguous bases on the basis of the following criteria: (a) scan position differences, (b) peak height ratios, (c) peak area ratios, and (d) base composition; and

evaluating the quality and coverage of the sample sequence information in comparison to the reference sequence information to identify reportable ranges and sequence variants for the sample sequence information.

2. The method of Claim 1, wherein the rule-based criteria for assessing scan position differences to differentiate between ambiguous and unambiguous bases further comprises identifying differences between scan positions of major and minor bases within the sample sequence information which fall below an empirical threshold.

3. The method of Claim 2, wherein the empirical threshold associated with identifying differences between scan positions is in the range of approximately 0 to approximately 3.

4. The method of Claim 1, wherein the rule-based criteria for assessing scan position differences for differentiating between ambiguous and unambiguous bases further comprises identifying differences between scan positions of major and

minor bases within the sample sequence information which reside above, below, or are substantially equivalent to a user-defined threshold.

5. The method of Claim 1, wherein the rule-based criteria for assessing peak height ratios to differentiate between ambiguous and unambiguous bases further comprises assessing peak height ratios for major and minor bases within the sample sequence information which exceed an empirical threshold.

6. The method of Claim 5, wherein the empirical threshold associated with assessing peak height ratios is in the range of approximately 0.3 to approximately 1.0.

7. The method of Claim 1, wherein the rule-based criteria for assessing peak height ratios for differentiating between ambiguous and unambiguous bases further comprises assessing peak area ratios of major and minor bases within the sample sequence information which reside above, below, or are substantially equivalent to a user-defined threshold.

8. The method of Claim 1, wherein the rule-based criteria for assessing peak area ratios to differentiate between ambiguous and unambiguous bases further comprises assessing peak area ratios for major and minor bases within the sample sequence information which exceed an empirical threshold.

9. The method of Claim 8, wherein the empirical threshold associated with assessing peak area ratios is in the range of approximately 0.3 to approximately 1.0.

10. The method of Claim 1, wherein the rule-based criteria for assessing peak area ratios for differentiating between ambiguous and unambiguous bases further comprises assessing peak area ratios of major and minor bases within the sample sequence information which reside above, below, or are substantially equivalent to a user-defined threshold.

11. The method of Claim 1, wherein the rule-based criteria for assessing base composition to differentiate between ambiguous and unambiguous bases further comprises determining if the major and minor bases within the sample sequence information are both purines or both pyrimidines.

12. The method of Claim 11, wherein when ambiguity in the base composition is increased when the major and minor bases are both purines or both pyrimidines.

13. The method of Claim 1, wherein the rule-based criteria for distinguishing between ambiguous and unambiguous bases comprises identifying consecutive runs of bases exceeding an empirical threshold.

14. The method of Claim 13, wherein the empirical threshold associated with comprises identifying consecutive runs of bases is in the range of approximately 10 to approximately 13.

15. The method of Claim 1, wherein the rule-based criteria for distinguishing between ambiguous and unambiguous bases identifies consecutive runs of bases which reside above, below, or are substantially equivalent to a user-defined threshold.

16. The method of Claim 1, wherein identified ambiguous bases are excluded from the evaluation of quality and coverage of the sample sequence information.

17. The method of Claim 1, wherein identified ambiguous bases are excluded from the identification of sequence variants.

18. The method of Claim 1, wherein the sequence information corresponds to mitochondrial DNA sequence information.

19. An automated method for mitochondrial DNA analysis used to identify associations between a target sample of unknown familial origin with that of at least one reference sample, the method comprising:

acquiring genetic information describing the sequence composition and characteristics for a plurality of nucleotides relating to the mitochondrial genetic makeup of the target sample and at least one reference sample;

assessing the genetic information to identify a degree of ambiguity associated with each of the plurality of nucleotides wherein ambiguous nucleotides are distinguished from unambiguous nucleotides on the basis of:  
(a) scan position differences, (b) peak height ratios, (c) peak area ratios, and  
(d) nucleotide compositions;

comparing the genetic information and the degree of ambiguity associated with each of the plurality of nucleotides of the target sample and the at least one reference sample to identify a nucleotide signature that provides distinguishing information used to identify sequence similarities and differences between the target sample and the at least one reference sample; and

comparing the nucleotide signature of the target sample to that of the at least one reference sample such that substantially identical nucleotide signatures identify the target sample as being of the same familial origin as the at least one reference sample and nucleotide signatures which are not substantially identical identify the target sample as not being of the same familial origin as the at least one reference sample.

20. The method of Claim 19, wherein the genetic information comprises, in part, raw sequence information selected from the group consisting of: electropherogram/sequence traces, information describing peak characteristics, peak height information, peak area information, peak width information, putative nucleotide identifications, base calls, quality value assessments, and scan positions.

21. The method of Claim 19 wherein, nucleotides associated with a threshold degree of ambiguity are excluded from inclusion in the nucleotide signature.

22. The method of Claim 19, wherein the genetic information relates to one or more hypervariable regions within the mitochondrial genome.

23. The method of Claim 19, wherein the origin of the genetic information for the target sample and/or the reference sample is skin, hair, salvia, semen, tissue, bone, or blood.

24. The method of Claim 19, wherein the genetic information for the reference sample comprises sequence information obtained from an original Cambridge reference sequence database or a revised Cambridge Reference Sequence database.

25. The method of Claim 19, wherein the use of scan position differences to differentiate between ambiguous and unambiguous nucleotides further comprises identifying differences between scan positions of major and minor bases within the sample sequence information which fall below an empirical threshold.

26. The method of Claim 19, wherein the use of scan position differences for differentiating between ambiguous and unambiguous nucleotides further comprises identifying differences between scan positions of major and minor bases within the sample sequence information which reside above, below, or are substantially equivalent to a user-defined threshold.

27. The method of Claim 19, wherein the use of peak height ratios to differentiate between ambiguous and unambiguous nucleotides further comprises

assessing peak height ratios for major and minor bases within the sample sequence information which exceed an empirical threshold.

28. The method of Claim 19, wherein the use of peak height ratios for differentiating between ambiguous and unambiguous nucleotides further comprises assessing peak area ratios of major and minor bases within the sample sequence information which reside above, below, or are substantially equivalent to a user-defined threshold.

29. The method of Claim 19, wherein the use of peak area ratios to differentiate between ambiguous and unambiguous nucleotides further comprises assessing peak area ratios for major and minor bases within the sample sequence information which exceed an empirical threshold.

30. The method of Claim 19, wherein the use of peak area ratios for differentiating between ambiguous and unambiguous nucleotides further comprises assessing peak area ratios of major and minor bases within the sample sequence information which reside above, below, or are substantially equivalent to a user-defined threshold.

31. The method of Claim 19, wherein the use of nucleotide composition to differentiate between ambiguous and unambiguous nucleotides further comprises determining if the major and minor bases within the sample sequence information are both purines or both pyrimidines.

32. The method of Claim 31, wherein when ambiguity in the nucleotide composition is increased when the major and minor bases are both purines or both pyrimidines.

33. The method of Claim 19, wherein ambiguous and unambiguous nucleotides are distinguished, in part, by identifying consecutive runs of nucleotides exceeding an empirical threshold.

34. The method of Claim 19, wherein ambiguous and unambiguous nucleotides are distinguished, in part, by identifying consecutive runs of nucleotides which reside above, below, or are substantially equivalent to a user-defined threshold.

35. A system for conducting automated comparison analyses of sequence information relating to at least one sample and at least one reference, the system comprising:

a setup module that acquires and formats sequence information relating to the at least one sample and the at least one reference;

a trace-analysis module that prepares the sequence information for comparison and includes functionality to select appropriate regions of the sequence information of the at least one sample and the at least one reference for subsequent comparison;

an assembly-analysis module that generates one or more consensus sequences between the at least one sample and the at least one reference and includes functionality for evaluating the sequence information to distinguish between ambiguous and unambiguous nucleotides within the sequence information on the basis of the following criteria: (a) scan position differences, (b) peak height ratios, (c) peak area ratios, and (d) nucleotide composition; and

a variant-analysis module that generates a nucleotide profile which details the results of comparing the at least one sample and the at least one reference and identifies nucleotide variations between the at least one sample and the at least one reference wherein the variants may be used to

determine the degree of similarity between the at least one sample and the at least one reference.

36. The system of Claim 35, wherein the use of scan position differences by the assembly-analysis module to differentiate between ambiguous and unambiguous nucleotides is conducted by identifying differences between scan positions of major and minor bases within the sample sequence information which fall below an empirical threshold.

37. The system of Claim 35, wherein the use of scan position differences by the assembly-analysis module to differentiate between ambiguous and unambiguous nucleotides further comprises identifying differences between scan positions of major and minor bases within the sample sequence information which reside above, below, or are substantially equivalent to a user-defined threshold.

38. The system of Claim 35, wherein the use of peak height ratios by the assembly-analysis module to differentiate between ambiguous and unambiguous nucleotides further comprises assessing peak height ratios for major and minor bases within the sample sequence information which exceed an empirical threshold.

39. The system of Claim 35, wherein the use of peak height ratios by the assembly-analysis module to differentiate between ambiguous and unambiguous nucleotides further comprises assessing peak area ratios of major and minor bases within the sample sequence information which reside above, below, or are substantially equivalent to a user-defined threshold.

40. The system of Claim 35, wherein the use of peak area ratios by the assembly-analysis module to differentiate between ambiguous and unambiguous nucleotides further comprises assessing peak area ratios for major and minor bases within the sample sequence information which exceed an empirical threshold.

41. The system of Claim 35, wherein the use of peak area ratios by the assembly-analysis module to differentiate between ambiguous and unambiguous nucleotides further comprises assessing peak area ratios of major and minor bases within the sample sequence information which reside above, below, or are substantially equivalent to a user-defined threshold.

42. The system of Claim 35, wherein the use of nucleotide composition by the assembly-analysis module to differentiate between ambiguous and unambiguous nucleotides further comprises determining if the major and minor bases within the sample sequence information are both purines or both pyrimidines.

43. A computer readable medium having stored thereon instructions which cause a general purpose computer to perform the steps of:

acquiring sequence information relating to at least one sample and to at least one reference for purposes of comparison;

evaluating the sequence information relating to the at least one sample to identify ambiguous bases present within the sample on the basis of the following criteria: (a) scan position differences, (b) peak height ratios, (c) peak area ratios, and (d) base composition; and

evaluating the quality and coverage of the sample sequence information in comparison to the reference sequence information to identify reportable ranges and sequence variants for the sample sequence information.

44. A computer-based system for performing automated sequence evaluation and used to identify associations between a target sample of unknown familial origin with that of at least one reference sample, the system comprising:

a database for storing genetic information describing the sequence composition and characteristics for a plurality of nucleotides relating to the

mitochondrial genetic makeup of the target sample and at least one reference sample;

a program which performs the operations of:

assessing the genetic information to identify a degree of ambiguity associated with each of the plurality of nucleotides wherein ambiguous nucleotides are distinguished from unambiguous nucleotides on the basis of:  
(a) scan position differences, (b) peak height ratios, (c) peak area ratios, and  
(d) nucleotide compositions;

comparing the genetic information and the degree of ambiguity associated with each of the plurality of nucleotides of the target sample and the at least one reference sample to identify a nucleotide signature that provides distinguishing information used to identify sequence similarities and differences between the target sample and the at least one reference sample;  
and

comparing the nucleotide signature of the target sample to that of the at least one reference sample such that substantially identical nucleotide signatures identify the target sample as being of the same familial origin as the at least one reference sample and nucleotide signatures which are not substantially identical identify the target sample as not being of the same familial origin as the at least one reference sample.